

P-value에 대한 고찰

P-value

P-value의 한계점

▶ 관련 기사 자료 및 논문 자료

<https://www.editage.co.kr/insights/is-my-research-si...> ▼

p-value(유의 확률)에 의존해서는 안 되는 이유 - 에디티지

📖 2016. 6. 2. — 1. **p-value**는 데이터가 특정 통계 모형과 얼마나 대립하는지 나타낼 수 있다. 여기서 중요한 단어는 “특정”입니다. 모든 연구나 분석에서 연구자는 통계 ...
이 페이지를 20. 12. 22에 방문했습니다.

<http://m.blog.naver.com/int9708> ▼

쉽게 빠지는 P-value의 함정 : 네이버 블로그

📖 2016. 10. 14. — 그런데 최근 생명과학과 의학, 사회과학, 심리학에서 많은 실험이 재현되지 않아 문제가 되고 있습니다. 그 원인으로 **P-value**가 지목되면서 몇몇 극단 ...

<https://niceguy1575.tistory.com/entry/가설검정에...> ▼

가설검정에서 P-value는 정말로 유효할까? :: 오늘도, 일희일비

📖 2017. 8. 2. — 이 **한계**를 악용하여 연구 성과를 입증하는 연구 사례도 많다고 합니다. 익히 알려진 **p-value**의 한계점은 다음과 같습니다. A. 관측치가 많을수록 **p**값이 ...

https://boxnwhis.kr/dont_be_overwhelmed_by_pvalue ▼

A/B 테스트에서 p-value에 휘둘리지 않기 :: -[]- Box and Whis...

📖 2016. 4. 15. — A/B 테스트에서 **p-value**에만 과하게 집중하는 것이 왜 좋지 않은지 설명 ... 즉, 미리 설정한 1종 오류의 허용 **한계**보다 이러한 데이터가 관찰될 확률 ...

<https://brunch.co.kr/...> ▼

통계, 기본 개념을 정리해보자. - 브런치

2019. 8. 24. — 2019 앱으로 보기. 이 포스트에서는 유의수준과 **p-value**(유의확률)이 무엇인지, 그 전에 귀무가설과 대립가설, 1종오류와 2종오류가 무엇인지 등을 정리한다. ... 유의수준을 0.05(5%)로 정했다면 0.05면 100번 실험을 했을 때 1종 오류를 범할 수 있는 최대 허용 **한계**를 5번으로 정하겠다는 말이다. 유의 수준은 보통 α (...

<http://scienceon.hani.co.kr/...> ▼

'통계적으로 유의미한 결과'는 얼마나 유의미할까 - 사이언스온

📖 2016. 5. 20. — 이와 같이 **p**값의 **한계**에 오해까지 겹쳐져, 유의성검정 절차는 과학 가설검정 절차에 대한 심각한 오해 및 오용의 원인이 되고 있다. **p**값을 이용한 가설 ...

▷ p-value의 한계에 대한 데이터 사이언스 기사와 사용시 유의점에 대한 글이 많음

P-value

P-value의 한계점

▶ P-value의 한계점 근거

- ▶ 1. p-value의 본래 목적 혼동: 귀무 가설이 맞다고 가정했을 때 극단적 데이터가 관측될 확률

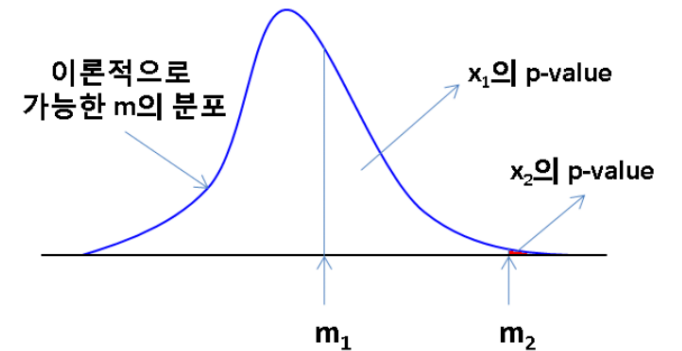
-> 즉, p-value는 단순히 대립 가설이 참일 확률이 아님

- ▶ 2. 관측치가 많을 수록 p-value 값이 작아진다.

-> 데이터의 수가 커진다면 C.I(신뢰구간)이 좁아지게 되고 이에 따라 p값도 유의하게 작아질 수 밖에 없다.

- ▶ 3. 선택적으로 p값과 통계적 방법론을 선택한다

-> 유의하지 않은 변수의 값을 제거하고 유의한 변수만을 채택하는 등 통계적 방법론을 조작하는 경우가 발생할 수 있다.



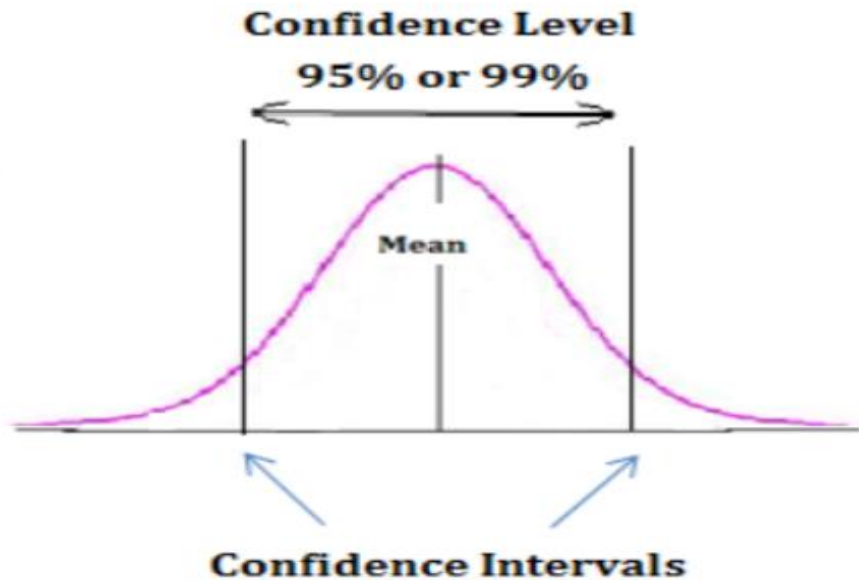
- 출처 : 위키피디아

▷ 즉, p-value 값만으로는 신뢰성을 얻기 힘들고 p-value 값과 더불어 신뢰성을 높일 자료를 함께 제시해야 한다.

P-value

P-value와 대안책

▶ 신뢰 구간



P-value와 신뢰구간

- ▶ 신뢰구간: 모수가 실제로 포함되어 있을 것으로 예상되는 범위
 - > 샘플링된 데이터를 기반으로 모수의 범위를 추정하기 위해 사용
 - > 공식을 이용해 confidence interval 값을 측정
- ▶ 만약 p값이 알파(유의 수준) 보다 낮다면 신뢰구간은 가설평균(hypothesized mean)을 포함하지 않음.

▷ 즉, p-value 값의 유의성을 따져본 후, 신뢰구간을 확인하면 좀 더 설득력 있는 정보를 얻을 수 있음

P-value

P-value외 대안책

▶ 효과 크기

Table 1. Illustrative Interpretations of Cohen's d

Estimated values	Proportion of control group which would be below the mean of the treatment group	Size of effect
0.0	50.0	Small effect
0.2	57.9	
0.4	65.5	
0.5	69.1	Medium effect
0.8	78.8	
1.2	88.5	Large effect
1.6	94.5	
2.0	97.7	
2.6	99.5	
3.0	99.9	

P-value와 효과크기

- ▶ 효과 크기: 실험에 사용 된 처치에 대한 효과의 크기를 표준화 추정된 통계량
-> 서로 다른 방법과 단위로 측정된 유사한 관측 값에 대한 통계 결과는 효과크기를 통하여 비교
-> 피어슨의 상관 계수 r , 오즈비(odds ratio, OR), Cohen's d
- ▶ Cohen's d : 표준정규분포에서 z 점수(z -score)와 비슷한 방법으로 계산
-> Cohen's d 는 비교하려는 두 군의 차이에 대한 백분율로 변환 가능.
-> 상관계수 r 로 쉽게 변환 가능

▷ p-value와 신뢰 구간은 유의성만을 검증, 정확한 실험에 대한 정보를 알기 어려움.
이를 해결하기 위해 효과크기와 같은 표준화 통계량을 함께 제시해야 함.

P-value

P-value와 대안책

▶ 베이지안 통계학 가설 검정

빈도 함수론

- ▶ 모집단에 대한 사전 정보를 사용하지 않음
- ▶ 모수치가 우연에 의한 값인지 아닌지를 통해 추정
- ▶ sample size가 클수록 좋음
- ▶ 영가설이 기각 되는지의 여부만 알 수 있음
- ▶ 상호 작용 효과를 볼 때, 집단 간 차이가 있다 없다만 알 수 있음

베이지안 통계 분석

- ▶ 모집단에 대한 사전 지식 사용
- ▶ 모수치는 애초에 불확실하다
- ▶ sample size에 크게 영향 받지 않음
- ▶ 영가설/대립가설이 어느 정도의 가능성으로 참인지 확률로써 알아낼 수 있음
- ▶ 집단 간 차이는 물론 유사성도 알 수 있음

▷ 사전 정보가 존재한다는 가정 하에 베이지안 통계 가설 검정을 이용하면 좀 더 명확하게 다양한 정보들을 이용해 집단 간 비교를 설명할 수 있음

통계적 검정을 통한 산업군 분석

통계적 검정을 통한 산업군 분석

신뢰구간

- ▶ 평균 비교를 위한 신뢰구간을 도출
- ▶ 업종 중 남성과 여성의 이용건수 평균에 차이가 있는지 산업마다 검정

가설 설정

$$H_0 : \mu_{19} = \mu_{20}$$

(남성과 여성의 택시 운송업 평균이용건수는 같다)

$$H_1 : \text{not } H_0$$

(남성과 여성의 택시 운송업 평균 이용건수는 다르다)

검정

- ▶ 유의수준 = 5%
- ▶ Scipy에서 제공하는 stats모듈 사용
- ▶ p-value 도출 및 차이가 있는 산업군 선정
- ▶ 신뢰구간 도출

통계적 검정을 통한 산업군 분석

신뢰구간

- ▶ 등분산일 경우와 이분산일 경우를 따름
- ▶ 유의수준 5%

등분산

$$s_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

자유도
DF = $n_1 + n_2 - 2$

$$s = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

이분산

$$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

자유도

$$DF = \frac{(\text{VAR}_1 + \text{VAR}_2)^2}{\frac{\text{VAR}_1^2}{n_1 - 1} + \frac{\text{VAR}_2^2}{n_2 - 1}}$$

신뢰구간

공식

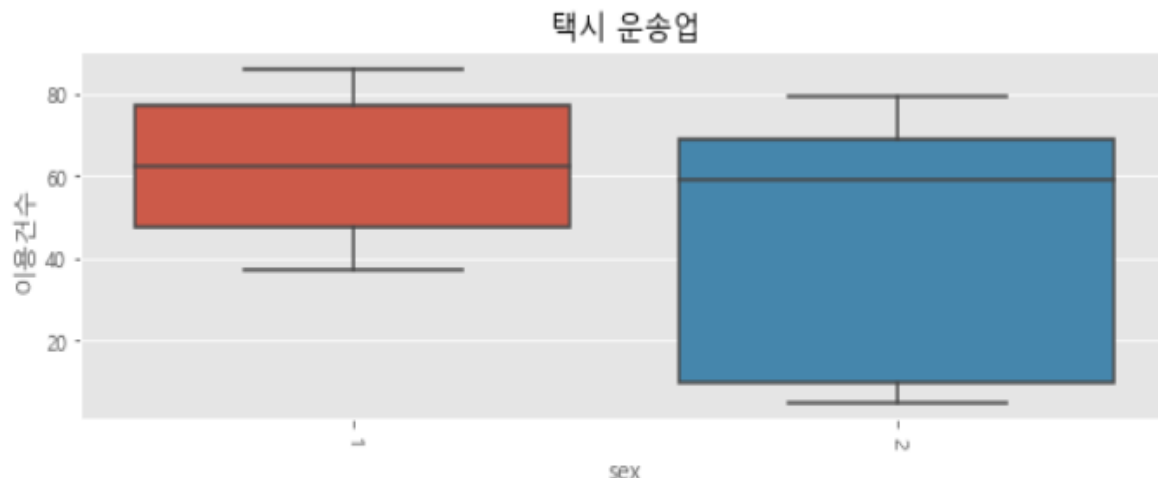
$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}(S) \sim (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}(S)$$

통계적 검정을 통한 산업군 분석

전라도 산업 검정

▶택시운송업

```
In [226]: sex_test(credit, "택시 운송업")
```



```
ShapiroResult(statistic=0.9593291878700256, pvalue=0.7746877670288086)
ShapiroResult(statistic=0.8466378450393677, pvalue=0.18412284553050995)
LeveneResult(statistic=0.4695489710590521, pvalue=0.5152376553991492)
택시 운송업 의 결과 : Ttest_indResult(statistic=0.8803553579901157, pvalue=0.4078733526479853)
16.7 < 17.6 < 18.5
```

- 남성 택시 운송업 이용건수 정규성 성립
- 여성 택시 운송업 이용건수 정규성 성립
- 남/여성 택시 운송업 이용건수 등분산성 성립

- ▶ 최종적으로 평균 차이가 없는 의견에 지지
- ▶ 하한 : 16.7
- ▶ 상한 : 18.5
- ▶ 남성 평균 - 여성 평균 = 17.6